Smolt Power Analysis

Report by:

Kathryn Irvine
University of Idaho, Moscow, Idaho.

Task Agreement: J8W07060004
Cooperative Agreement: H8W07060001

Submitted to the San Francisco Bay Area Network, Golden Gate National Recreation Area, Sausalito, CA 94965.

May 2008

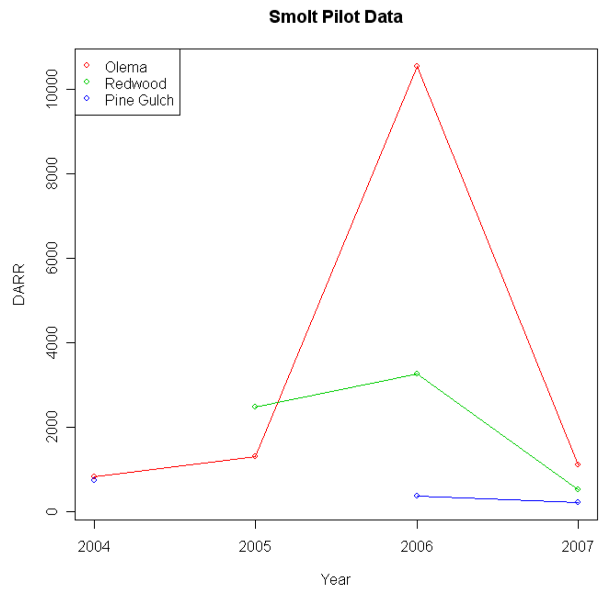
This report summarizes the power analysis for smolt trapping conducted by the SFAN at Olema, Redwood, and Pine Gulch Creeks. The power analysis is based on the pilot data provided by Michael Reichmuth.

DATA:

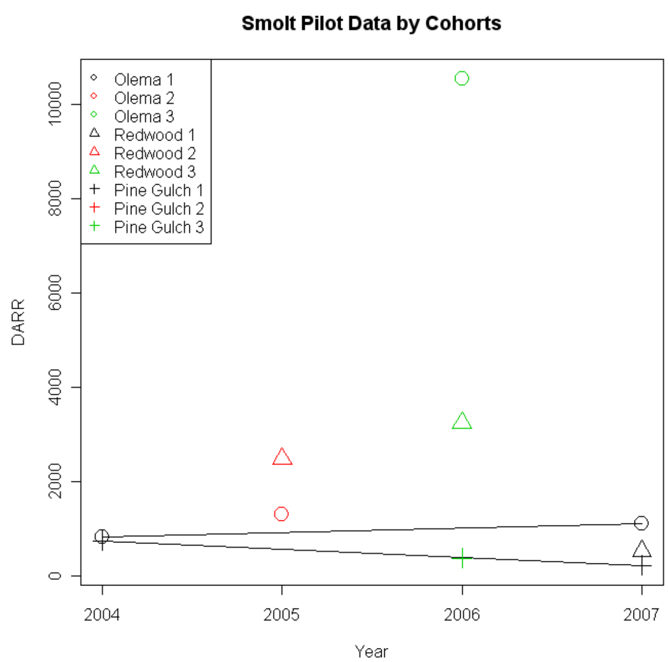
I used the data in Table 1 for the power analysis. The response used was the DARR estimates provided by the SFAN (Bjorkstedt 2000). Notice there is really only one year of data for each cohort. The numbering of the cohorts is arbitrary, but I selected 2005 as cohort year 1 because Olema and Redwood creeks had data for 2005. In the power analysis I just assumed the DARR estimate for Pine Gulch creek was 737, the estimate from 2004. Because there is only one observation for each cohort, we are unable to pursue an approach that utilizes a mixed linear model for trend because we do not have preliminary estimates for some of the variance components involving time in such a model.

Table 1. Data used for smolt power analysis

Creek	Year	Species	Age_Class	total_catch	DARR	Cohort
Olema	2004	Coho	Smolt	229	831	3
Olema	2005	Coho	Smolt	87	1296	1
Olema	2006	Coho	smolt	368	10544	2
Olema	2007	Coho	smolt	330	1098	3
Redwood	2005	Coho	smolt	301	2481	1
Redwood	2006	Coho	smolt	1048	3253	2
Redwood	2007	Coho	smolt	330	520	3
Pine Gulch	1999	Coho	smolt	0	NA	1
Pine Gulch	2002	Coho	smolt	249	NA	1
Pine Gulch	2003	Coho	smolt	576	NA	2
Pine Gulch	2004	Coho	smolt	149	737	3
Pine Gulch	2005	Coho	smolt	8	NA	1
Pine Gulch	2006	Coho	smolt	93	368	2
Pine Gulch	2007	Coho	smolt	76	219	3



The above figure displays the data from the Table 1. The lines are connecting the observations for each creek without distinguishing the three different cohorts.



The above figure distinguishes the three different cohorts for each creek. The observations for each creek have the same symbol and the different colors represent the there different cohorts (black= cohort 1, red= cohort 2, green= cohort 3). The lines connect the observations for the same cohorts at Olema and Pine Gulch Creeks. Only cohort 1 at Olema and Pine Gulch Creeks had two observations in 2004 and 2007. To simplify the data for the power analysis, I only considered 2005, 2006, and 2007.

POWER ANALYSIS:

In order to perform a power analysis for trend, a model must be assumed for the future data. I assume a linear model that specifies a creek and cohort fixed effect because these will be the same throughout the study. The power based on this simplified model can be determined using PROC GLMPOWER in SAS.

The following linear model for the log of DARR estimates was used:

$$ln(y_{ijk}) = \gamma_k + \alpha_i + w_j\beta + d_{ik} \qquad (1)$$

where $i = 1,\ldots, m_a$; $j = 1, \ldots, m_b$; $k = 1,\ldots,m_c$ (i indexes the creek, j indexes the year of sampling, and k indexes the cohort). Let

m_a = the number of creeks surveyed for smolts ($m_a = 3$);
m_b = the number of years of sampling;
m_c= the number of cohorts ($m_c = 3$);
y_{ijk} = DARR estimate for smolts at creek i, year j, and cohort k.
w_j = is a constant representing the jth year (covariate);
β = fixed slope of the linear time trend;
α_i = the fixed intercept of the i^{th} creek;
γ_k =fixed intercept for the k^{th} cohort, iid as $N(0, \sigma^2_b)$; and
d_{ik} = the random error (creek*cohort interaction term), iid as $N(0, \sigma^2_d)$.

The common linear trend to all creeks and cohorts is denoted by β. For the power analysis, we are assuming different values for β and then determining the probability of detecting that level of trend for a specified number of sampling years. This overall trend is then modified for each creek and cohort by the intercept terms, α_i and γ_k respectively.

To test for a linear trend in the medians we are interested in testing the null hypothesis of $\beta=0$ (no trend) versus the alternative hypothesis of $\beta \neq 0$ (three-year trend). As defined above, β is the coefficient of the Year variable (referred to as WYear in the SAS code) and represents the three-year trend of the outcome on the natural log scale. The trend is coded as every three years because of the biology of smolts, the same cohort is observed every three years

The three-year trend on the logged scale is estimated by $\hat{\beta}$ the coefficient of the Year variable. The three-year trend on the original scale is estimated by $\exp(\hat{\beta})$. If we are interested in interpreting the percent change in the medians of the original scale we need to back-transform the confidence interval end points. Confidence intervals for the three-year percent change, $\exp(\hat{\beta})-1$ are found by applying the exponential function to the endpoints of the confidence interval for $\hat{\beta}$ and subtracting 1. In other words, if the (1-α)100 % confidence interval for β is (a,b), then the (1-α)100% confidence interval for the percent three-year change in the medians is (exp(a)-1,exp(b)-1). For example, if the three-year trend on the natural log scale is estimated as -0.25 with 95% confidence interval between -0.5 and -0.15. We would state for the original scale: We are 95% confident the median DARR smolt estimates are decreasing every three years between -39% to -14 % (-39%, -14% 95% confidence interval).

For completeness, we should mention an alternative approach that could be pursued if more pilot data were available. A linear mixed model, similar to Piepho and Ogutu (2002), could be used if more years of pilot data were available. The trend model for smolts would be altered to reflect the three different creeks and cohorts. At this point we are unable to estimate any of the random effects in such a mixed model because we do not have information about the potential variability in trends between cohorts and creeks for the power analysis. The power based on the simplified model (Model 1) will likely be over-estimated since we are assuming less variability in the data.

METHODS:

To calculate the power for α=.10, three creeks, and three cohorts we use PROC GLMPOWER in SAS. We specify two levels of decline a 2.5% three-year decrease for each cohort and a 10% three-year decrease for each cohort. To generate data under these two different decline scenarios the observed data for each creek and cohort are used as the baseline year (Y_0). The data is then projected forward in time by $Y_t = bY_{t-1}$ where b is either .90 (10% three-year decline) or .975 (2.5% three-year decline) and t indexes time. The Wyear variable is the cohort year of sampling. For example, 0 for the first year of sampling cohort 1, 2, or 3 and then 3 for the second time cohort 1, 2 or 3 is sampled.

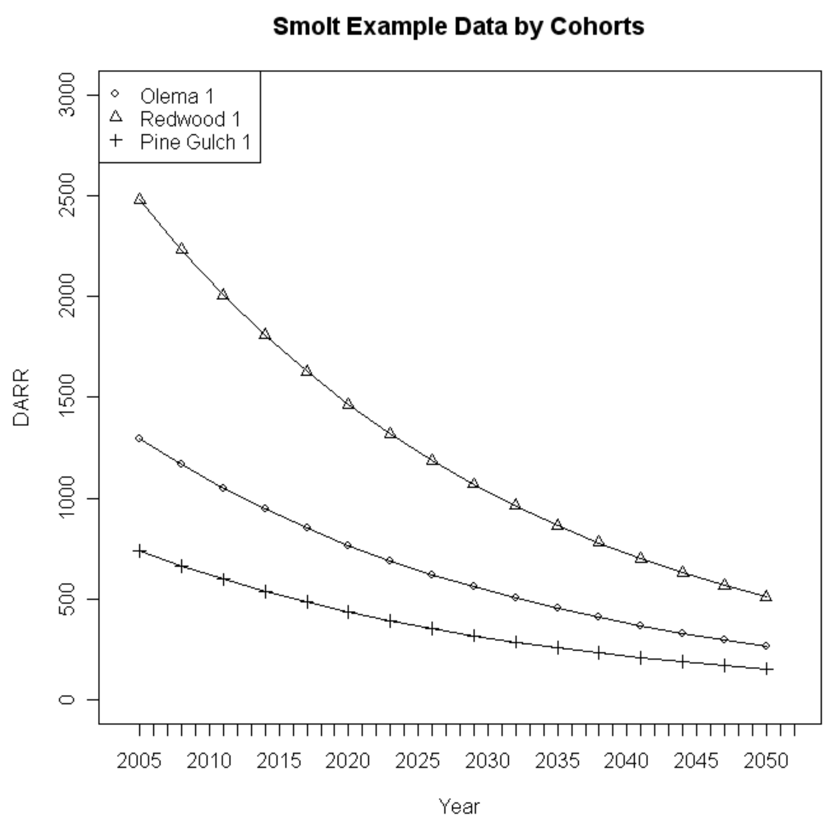
```
EXAMPLE SAS CODE:
proc glmpower data=year8dec25;
  class Cohort Creek WYear;
  model lnDARR = Creek Cohort WYear;
 contrast 'linear' WYear -.54 -.386 -.231 -.077 .077 .231 .386 .54 ; /*
by sampling interval by cohorts*/
  power
    alpha=.10
    stddev = .748 /*THIS ESTIMATE corresponds to the creek*cohort term
in previous model*/
      Ntotal = 72
      power = .;
run;
```

RESULTS:

The following table summarizes the power for α=.10, three creeks, and two declines (2.5% and 10% every three years) for different lengths of sampling.

Power	Year	% decline every 3 years
.102	2005-2016 (4 per cohort)	2.5
.114	2005-2028 (8 per cohort)	2.5
.146	2005-2040 (12 per cohort)	2.5
.454	2005-2052 (16 per cohort)	2.5
.127	2005-2016 (4 per cohort)	10
.322	2005-2028 (8 per cohort)	10
.703	2005-2040 (12 per cohort)	10
.957	2005-2052 (16 per cohort)	10

Based on the power analysis, for the proposed sampling design of smolts (Olema, Redwood, and Pine Gulch Creeks) to achieve 80% power we would need to have greater than 12 observations for each cohort at all three creeks in order to detect a 10% three-year decline. Because of the three-year return cycle for smolts this would mean greater than 35 years of sampling annually. The following figure displays an example of a three-year 10% decline for cohort 1 at all three creeks.



REFERENCES:

Piepho, H-P and J.O. Ogutu. 2002. A simple mixed model for trend analysis in wildlife populations. JABES 7: 350-360.

Bjorkstedt, E. P. 2000. DARR (Darroch Analysis with Rank-Reduction): A method for analysis of stratified mark-recapture data from small populations, with application to estimating abundance of smolts from outmigrant trap data. Administrative Report SC-00-02. Santa Cruz Laboratory Contribution Number 116.